Metode *Decison Tree C4.5* untuk Klasifikasi penyakit Jantung

Tutuk Indriyani[1], Fajar Fahru Rozi[2], Maftahatul Hakimah[3], Nanang Fakhrur Rozi[4] dan Rani Rotul Muhima[5]

Institut Teknologi Adhi Tama Surabaya[1,2,3,4,5]
e-mail: tutuk@itats.ac.id

ABSTRACT

Various types of machine learning studies have been conducted, one of which is a study on the use of machine learning to predict heart disease. In this study, 304 datasets were used for classification. Based on the background and results of previous studies, the author decided to classify heart disease using the Decison Tree C4.5 method. The algorithm classified correctly when the test results using the confusion matrix showed an Accuracy value of 0.86, which indicated that the classification of this test dataset was 86% and 14% were not classified correctly. Overall, the algorithm classified the heart disease dataset well. This is indicated by an average value such as precision of 0.87, which means that of all the predictions made by the model, around 87% were correct, then the recall result was 0.84, which means that the model successfully detected around 84%.

Kata kunci: *classification, Decision tree C4.5, heart disease, machine learning.*

ABSTRAK

Berbagai macam penelitian *machine learning* sudah dilakukan, salah satunya merupakan penelitian pemanfaatan machine learning untuk memprediksi penyakit jantung. Dalam Penelitian ini 304 dataset digunakan untuk melakukan klasifikasi. berdasarkan latar belakang dan hasil penelitian sebelumnya, penulis memutuskan untuk mengklasifikasi penyakit jantung menggunakan metode *Decison Tree* C4.5 Algoritma mengklasifikasi dengan benar saat hasil uji menggunakan confusion matrix menunjukkan nilai *Accuracy* 0,86 yang menunjukkan klasifikasi dataset pengujian ini sebesar 86% dan 14% tidak terklasifikasi dengan benar. Secara kesuluruhan, algoritma mengklasifikasi dataset penyakit jantung dengan baik. Hal ini di indikasikan dengan nilai rata rata seperti precision 0,87 yang berarti dari semua prediksi yang model dibuat sekitar 87% adalah benar, kemudian hasil *recall* adalah 0.84, yang berarti model berhasil mendeteksi sekitar 84%.

Kata kunci: klasifikasi, *Decision tree C4.5,* penyakit jantung, *machine learning.*

PENDAHULUAN

Menjaga kesehatan jantung sangat penting bagi kelangsungan hidup manusia karena jantung mendukung semua jaringan tubuh dan memainkan peran penting dalam aliran darah. Jantung merupakan salah satu organ tubuh yang berperan penting bagi manusia. Sebagai pengendali utama dalam sistem peredaran darah, jantung bekerja tanpa mengenal lelah untuk memompa darah. Di seluruh tubuh, terjadi proses esensial yang disebut kontraksi dan relaksasi teratur, di mana otot jantung memompa darah yang sarat dengan oksigen dari paru-paru (relaksasi). Itu juga memompa darah yang mengandung sisa metabolisme dan karbon dioksida dari semua bagian tubuh lainnya. Kontraksi jantung mengangkut darah ini ke paru-paru di mana ia ditukar dengan oksigen (kontraksi). Pada dasarnya, sangat penting bagi jantung untuk berada dalam kondisi prima untuk memastikan bagian tubuh lainnya berfungsi dengan lancar.

Penyakit jantung di Indonesia sedang meningkat dan tidak terbatas pada kelompok usia tertentu. Studi RISKESDAS (Riset Kesehatan Dasar) yang dilakukan pada 2020 melaporkan bahwa penyakit jantung menyerang 20 dari 1.000 orang Indonesia. Lebih lanjut, studi SRS *(Survei Sample Registration System)* tahun 2014 mengungkapkan bahwa penyakit jantung bertanggung jawab atas sekitar 12,9% kematian di Indonesia. Penyakit yang mengkhawatirkan ini menuntut peningkatan kewaspadaan dalam mendeteksi dan mencegah timbulnya kondisi ini. Banyak masyarakat yang masih awam terhadap kesehatan jantung sehingga mereka tidak sadar bahwa mereka menderita penyakit jantung. Hal tersebut dikarenakan pemeriksaan medis tentang kesehatan jantung yang

jarang dilakukan. Padahal penyakit ini bisa menyerang siapa saja bahkan seseorang yang tidak memiliki riwayat penyakit sebelumnya.

Banyak tugas *Machine Learning* menargetkan masalah klasifikasi. Berbagai macam penelitian *machine learning* sudah dilakukan, salah satunya merupakan penelitian pemanfaatan machine learning untuk memprediksi penyakit jantung. *Machine learning* adalah pembelajaran mesin yg sangat membantu pada penyelesaian masalah, praktis dalam mengerjakan sesuatu. Dibidang kesehatan, machine learning sangat praktis dalam mengerjakan sesuatu, contohnya dokter mampu mendiagnosa penyakit jantung secara singkat tanpa memakan waktu yang lama.

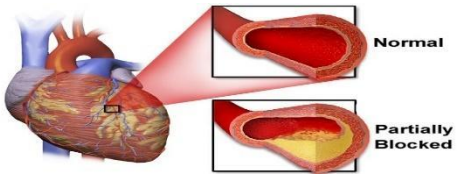
Terdapat beberapa metode dalam melakukan klasifikasi, antara lain *Decision Tree*. *Decision tree* adalah struktur *flowchart* yang mempunyai *tree* (pohon), dimana setiap simpul *internal* menandakan suatu tes atribut, setiap cabang merepresentasikan yang akan terjadi tes, dan simpul daun merepresentasikan kelas atau distribusi kelas. Alur di *decision tree* ditelusuri berasal simpul ke akar ke simpul daun yang memegang prediksi kelas. *Decision tree* adalah salah satu metode yang dipergunakan untuk pengklasifikasian dan prediksi karena memiliki kemudahan dalam interpretasi hasil [1]. Kelebihan algoritma *Decision Tree* adalah eliminasi perhitungan-perhitungan yang tak diharapkan, karena waktu memakai metode *decision tree* maka sample diuji hanya berdasarkan kriteria atau kelas tertentu.

Penelitian penelitian sebelumnya yang berkaitan meliputi : Klasifikasi Penderita Penyakit Diabetes Menggunakan Algoritma *Decision Tree* C4.5", "Analisis data hasil diagnosa untuk klasifikasi gangguan kepribadian menggunakan algoritma c4.5" , "analisa prediksi mahasiswa *drop out* menggunakan metode decision tree dengan algoritma id3 dan c4.5", dengan hasil pengukuran akurasi menggunakan ID3 diperoleh rata-rata 95,17%, sedangkan algoritma *Decision Tree* C4.5 diperoleh rata-rata akurasi sebesar 96,45% dan "Perbandingan Algoritma C4.5 Dan Id3 Untuk Prediksi Ketepatan Waktu Lulus Mahasiswa", dengan hasil pengukuran akurasi menggunakan ID3 diperoleh sebesar 78,75%, sedangkan algoritma *Decision Tree C4.5* diperoleh akurasi sebesar 81,88%.

Dalam penelitian ini, penulis mengklasifikasi penyakit jantung yang di ambil dari kaggle. Penulis akan mencari informasi dan mengklasifikasi apakah seseorang memiliki penyakit jantung atau tidak. Hal ini akan membantu tenaga medis dengan memberikan informasi seseorang memiliki penyakit jantung atau tidak. Dalam Penelitian ini 304 dataset digunakan untuk melakukan klasifikasi. berdasarkan latar belakang dan hasil penelitian sebelumnya, penulis memutuskan untuk mengklasifikasi penyakit jantung menggunakan metode *Decison Tree* C4.5

TINJAUAN PUSTAKA

Penyakit jantung adalah istilah umum untuk semua jenis gangguan yg mempengaruhi jantung. Penyakit jantung berarti sama dengan penyakit jantung namun tidak penyakit kardiovaskular. Penyakit kardiovaskular mengacu pada gangguan pembuluh darah serta jantung, sedangkan penyakit jantung mengacu hanya hati [2]. Penyakit jantung merupakan masalah kesehatan utama yang mepengaruhi salah satu organ vital tubuh. Meski memainkan peran penting, jantung adalah salah satu organ yang paling rentan terhadap penyakit [3].



Gambar 1. Penyakit Jantung Koroner (Blausen, 2014)

Decision Tree

Metode *decission tree* adalah sebuah struktur flowchart yang mirip struktur pohon, setiap titik pohon adalah atribut yang telah diuji, setiap cabang artinya hasil uji dan titik akhir merupakan pembagian kelas yg dihasilkan (Han dan Kamber, 2001). Pohon (*tree*) merupakan sebuah struktur

data yang terdiri dari simpul *(node*) serta rusuk (*edge*). Simpul di sebuah pohon dibedakan sebagai tiga, yaitu simpul akar (*root/node*), simpul percabangan/internal (*branch/internal node*) dan simpul daun (*leaf node*). Pohon keputusan merupakan representasi sederhana dari teknik klasifikasi untuk sejumlah kelas, dimana simpul internal juga simpul akar ditandai dengan nama atribut, setiap rusuknya diberi label nilai atribut yang mungkin dan simpul daun ditandai dengan kelas-kelas yang berbeda [6].

***Decision Tree* C4.5**

Algoritma C4.5 merupakan sebuah algoritma yang digunakan untuk membuat decision tree (pohon keputusan). Pohon keputusan adalah metode klasifikasi dan prediksi yang sangat bertenaga serta populer. algoritma C4.5 membangun pohon keputusan dari data pelatihan yang berupa kasus-kasus atau record-record (tupel) pada basis data. Setiap kasus berisikan nilai asal atribut-atribut buat sebuah kelas. Setiap atribut bisa berisi data diskret atau kontinyu (numerik). algoritma C4.5 juga menangani kasus yang tidak mempunyai nilai buat sebuah atau lebih atribut. Tapi, atribut kelas hanya bertipe data diskret serta tidak boleh kosong. Secara umum algoritma C4.5 untuk menciptakan pohon keputusan sebagai berikut :

1. Pilih atribut menjadi akar
2. Buat cabang buat masing-masing nilai
3. Bagi kasus pada cabang
4. Ulangi proses buat masing-masing cabang sampai semua kasus di cabang mempunyai kelas yang sama.

Entropy* dan *Gain

Sebuah obyek yang diklasifikasikan pada pohon wajib dites nilai *Entropy*nya. *Entropy* artinya ukuran berasal dari teori informasi yang dapat mengetahui karakteristik yang berasal *impuryt* dan *homogenity* dari kumpulan data. Nilai *Entropy* tersebut kemudian dihitung nilai gain masing-masing atribut. *Entropy* (S) adalah jumlah bit yang bisa mengekstrak suatu kelas (+ atau -) dari sejumlah data acak di ruang sampel S. *Entropy* bisa diartikan sebagai kebutuhan bit untuk menyatakan suatu kelas. Semakin kecil nilai entropy maka semakin *entropy* dipergunakan dalam mengekstrak suatu kelas. *Entropy* dipergunakan untuk mengukur ketidakaslian S.sistem informasi atau disebut processing system. Perhitungan nilai *entropy* menggunakan rumus berikut:

Entropy (S) = $\sum_{i=1}^{n} \quad -pi*pi$ .....(1)

Rumus menghitung nilai entropy

Keterangan :

S : himpunan kasus
A : fitur
n : Jumlah partisi S
pi : proporsi dari Si terhadap S

Gain merupakan salah satu *attribute selection measure* yang digunakan untuk menentukan test attribute tiap node pada tree. Atribut dengan *gain* tertinggi dipilih menjadi test atribut dari suatu node. *Gain* (S,A) artinya perolehan informasi dari atribut A relatif terhadap hasil data S. Perolehan informasi didapat berasal dari *output* data atau *variable dependent* S yang dikelompokkan berdasarkan atribut A, dinotasikan menggunakan *gain* (S,A) (Larissa Navia Rani, 2016). Untuk menghitung *gain* digunakan rumus berikut:

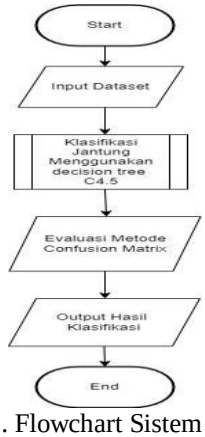
Gain (S,A) = Entropy (S) - $\sum_{i=1}^{n} \quad *$ Entropi(Si) .....(2)

Keterangan :

S : himpunan kasus
A : atribut
n : jumlah partisi artribut A
|Si| : jumlah kasus pada partisi ke-i
|S| : jumlah kasus dalam S

METODE

Proses tahapan-tahapan dalam penelitian ini dapat di tunjukkan pada Gambar 2.



Gambar 2. Flowchart Sistem Penelitian

Dari Gambar 2, maka tahapan yang dilakukan dalam penelitian ini adalah sebagai berikut:
Pengumpulan dataset. Data di klasifikasi untuk menemukan potensi penyakit jantung berdasarkan parameter yang ada dengan dengan menerapkan algoritma *decision tree* C4.5. Melakukan pengujian menggunakan *confuison matrix*[8]. Menampilkan output dari hasil klasifikasi.

Input Dataset

Dataset penyakit jantung yang digunakan dalam penelitian ini di ambil dari https://www.kaggle.com/code/dirgahalimsusilo/klasifikasi-penyakit-jantung/data. Dataset ini mencakup informasi tentang parameter-parameter seperti umur (*age*), kelamin (*sex*), tipe nyeri dada (*cp*), tekanan darah saat istirahat (*trestbps*), koresterol (*chol*), gula darah saat puasa (*fbs*), tes penyakit jantung (*restecg*), detak jantung (*thalach*), nyeri dada saat olahraga (*exang*), penurunan serangan jantung (*oldpeak*), puncak serangan jantung (*slope*), jumlah pembuluh darah (*ca*), hasil uji stress penyakit jantung (*thal*). Jumlah data yang tersedia dalam dataset ini adalah 303.

Klasifikasi Menggunakan Algoritma *Decision Tree* C4.5

Tahapan proses *Decision Tree* C4.5 di jelaskan pada Gambar 3.2 :



Gambar 3 Flowchart Algoritma *Decision Tree* C4.5

Langkah-langkah atau alur dalam penyelesaian perhitungan manual data mining menggunakan *Decision Tree* C4.5 sebagai berikut:

Input semua atribut, umur (*age*), kelamin (*sex*), tipe nyeri dada (*cp*), tekanan darah saat istirahat (*trestbps*), koresterol (*chol*), gula darah saat puasa (*fbs*), tes penyakit jantung (*restecg*), detak jantung (*thalach*), nyeri dada saat olahraga (*exang*), penurunan serangan jantung (*oldpeak*), puncak serangan jantung (*slope*), jumlah pembuluh darah (*ca*), hasil uji stress penyakit jantung (*thal*). Menghitung jumlah kasus yang yang di ambil dari tabel "target" yaitu total ada 14 kasus, yang terdiri dari 10 tidak dan 4 ya. Menghitung *entropy* dari "total" jumlah kasus. Menghitung *entropy* total menggunakan rumus(2).

Entropy (total) = (- 10/14*(10/14))+(- 4/14*(4/14))

= 0,863121

Pengujian Menggunakan *Confusion Matrix*

Melakukan pengujian menggunakan *confision matrix*. Pada pengukuran kinerja menggunakan *confusion matrix*, terdapat 4 istilah sebagai tersebut adalah *True Positive* (TP), *True Negative* (TN), *False Negative* (FN). Nilai *True Negative* (TN) merupakan data negatif yang terdeteksi benar kelas negatif, sedangkan *False Negative* (FN) merupakan kebalikan dari *False Positive*, sehingga data negatif, namun terdeteksi sebagai data positif. Sementara itu, *True Positive* (TP) merupakan data positif yang terdeteksi benar kelas positif, sedangkan *False Positive* (FP) merupakan kebalikan dari *False Negative*, sehingga data postif, namun terdeteksi sebagai data negatif. Berikut cara untuk menghitung *confusion matrix*[9] [10]:

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} * 100\%$$

$$\text{Presisi} = \frac{TP}{FP+TP} * 100\%$$

$$\text{Recall} = \frac{TP}{FN+TP} * 100\%$$

$$\text{F1-score} = 2 * \frac{(recall*presisi)}{(recall+presisi)} * 100\%$$

HASIL DAN PEMBAHASAN

Pembahasan Data I

Menampilkan data *training* pengujian untuk melakukan klasifikasi

Tabel 1. Data *Training*

no	no. acak	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal
1	132	42	1	1	120	295	0	1	162	0	0	2	0	2
2	202	58	1	0	150	270	0	0	111	1	0.8	2	0	3
3	196	46	1	2	150	231	0	1	147	0	3.6	1	0	2
4	75	55	0	1	135	250	0	0	161	0	1.4	1	0	2
5	176	60	1	0	117	230	1	1	160	1	1.4	2	2	3
..	..	..	..	..	..	..	..	..	..	..	..	..	..	..

238	188	50	1	2	140	233	0	1	163	0	0.6	1	1	3
239	71	51	1	2	94	227	0	1	154	1	0	2	1	3
240	106	69	1	3	160	234	1	0	131	0	0.1	1	1	2
241	270	46	1	0	120	249	0	0	144	0	0.8	2	0	3
242	102	63	0	1	140	195	0	1	179	0	0	2	2	2

Dari Tabel 1. diatas yang di tampilkan di program hanya *head* dan *tail* nya saja yaitu 5 teratas dan 5 terbawah. Data *training* adalah fase pada data yang digunakan untuk membangun sebuah model. Data *training* yang berasal dari file *.csv*, kemudian dataset berhasil dibagi ke data training, maka program akan menampilakan sebuah tabel yang berisi data training sebesar 80%, sehingga didapat sebanyak 242 data *training* dan data yang dihasilkan dari program urutanya acak.

Data *Testing*

Menampilkan data *testing* pengujian dari program untuk melakukan klasifikasi seperti berikut:

Tabel 2. Data *testing*

no	no. acak	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal
1	179	57	1	0	150	276	0	0	112	1	0.6	1	1	1
2	228	59	1	3	170	288	0	0	159	0	0.2	1	0	3
3	111	57	1	2	150	126	1	1	173	0	0.2	2	1	3
4	246	56	0	0	134	409	0	0	150	1	1.9	1	2	3
5	60	71	0	2	110	265	1	0	130	0	0	2	1	2
..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
57	249	69	1	2	140	254	0	0	146	0	2	1	3	3
58	104	50	1	2	129	196	0	1	163	0	0	2	0	2
59	300	68	1	0	144	193	1	1	141	0	3.4	1	2	3
60	193	60	1	0	145	282	0	0	142	1	2.8	1	2	3
61	184	50	1	0	150	243	0	0	128	0	2.6	1	0	3

Dari Tabel 2. diatas yang di tampilkan di program hanya *head* dan *tail* nya saja yaitu 5 teratas dan 5 terbawah. Proses *testing* merupakan sebuah proses pengujian dimana data yang data yang tidak diketahui kelasnya dimasukkan kedalam program yang telah di bentuk pada saat proses *training*. Pada saat proses data testing ini juga dilakukan dengan menggunakan data yang berasal dari suatu file yang formatnya *.csv*, lalu setelah dataset berhasil dibagi ke data *testing*, maka program akan menampilakan sebuah tabel yang berisi data *testing* sebesar 20%.

Informasi Node dan *Entropy*

Informasi Node dengan menerapkan rumus *entropy*

Tabel 3. Tabel *Node Decision Tree C4.5*

No	Feature	Threshold	Entropy	Gain
1	cp	0.5	0.992894	0.164763
2	ca	0.5	0.869893	0.18776
3	oldpeak	1.95	0.786368	0.10233

4	exang	0.5	0.997688	0.162368
5	exang	0.5	0.366578	0.041567
6	thal	2.5	0.69129	0.055477
7	oldpeak	0.7	0.77935	0.122433
8	age	44.5	0.684038	0.89403
9	slope	0.5	0.684038	0.684038
10	sex	0.5	0.789271	0.126404
11	ca	0.5	0.384312	0.02133
12	trestbps	115	0.322757	0.322757
13	trestbps	138	0.811278	0.125803
14	exang	0.5	0.970951	0.970951
15	exang	0.5	0.970851	0.970751

Feature: Ini adalah fitur yang digunakan oleh pohon keputusan untuk membagi data di simpul tersebut. *Threshold*: Nilai ambang batas yang digunakan oleh pohon keputusan untuk memisahkan data berdasarkan fitur yang terkait. *Impurity (Entropy)*: Semakin tinggi nilai *Impurity (Entropy)*, semakin tidak pasti dan campur aduk data pada simpul tersebut. *Gain*: Atribut dengan *gain* tertinggi dipilih menjadi tes atribut dari suatu node.

Hasil Klasifikasi

Menampilkan hasil klasifikasi pada data testing dari program seperti berikut:

Tabel 4 Hasil Klasifikasi

no	no. acak	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target	predicted
1	179	57	1	0	150	276	0	0	112	1	0.6	1	1	1	0	0
2	228	59	1	3	170	288	0	0	159	0	0.2	1	0	3	0	1
3	111	57	1	2	150	126	1	1	173	0	0.2	2	1	3	1	1
4	246	56	0	0	134	409	0	0	150	1	1.9	1	2	3	0	0
5	60	71	0	2	110	265	1	0	130	0	0	2	1	2	1	0
..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
57	249	69	1	2	140	254	0	0	146	0	2	1	3	3	0	0
58	104	50	1	2	129	196	0	1	163	0	0	2	0	2	1	1
59	300	68	1	0	144	193	1	1	141	0	3.4	1	2	3	0	0
60	193	60	1	0	145	282	0	0	142	1	2.8	1	2	3	0	0
61	184	50	1	0	150	243	0	0	128	0	2.6	1	0	3	0	0

Dari tabel 4 diatas yang di tampilkan di program hanya *head* dan *tail* nya saja yaitu 5 teratas dan 5 terbawah. Pada saat proses data testing ini juga dilakukan dengan menggunakan data yang berasal dari suatu file yang formatnya *.csv*. Proses klasifikasi pada data testing menghasilkan data yang berbeda, misalnya pada data nomor 228 pasien sebelumnya tidak memiliki penyakit jantung di hasil (target) tapi setelah di klasifikasi ternyata pasien menunjukkan angka 1 di *predicted* yang berarti pasien nomor 228 memiliki penyakit jantung dan seterusnya. Kemudian kita akan menguji hasil klasifikasi menggunakan metode *confusion matrix.*

Confusion Matrix

Hasil implementasi algoritma dapat dilihat melalui *confusion matrix*, dapat diketahui dari 61 data terdapat 31 data yang memiliki penyakit jantung dan 30 data yang tidak memiliki penyakit jantung. Sehingga terdapat 61 data testing yang terdiri dari TP = 27, FP = 4 , TN = 25, FN = 5.

$$\text{Accuracy} = \frac{(27+25)}{(27+5+4+25)} = 0{,}85$$

$$\text{Precision} = \frac{27}{(4+27)} = 0{,}87$$

$$\text{Recall} = \frac{27}{(5+27)} = 0{,}84$$

$$\text{F1_Score} = 2 * \frac{(0{,}84*0{,}87)}{(0{,}84+0{,}87)} = 0{,}86$$

Hasil dari *Classification Report* tersebut memberikan evaluasi kinerja model klasifikasi Anda berdasarkan beberapa metrik. Berikut adalah cara membaca setiap bagian dari laporan tersebut: *Precision*: Ini adalah ukuran sejauh mana prediksi positif yang dibuat oleh model adalah benar. Dalam hal ini, precision adalah 0.87, yang berarti dari semua prediksi yang model buat sekitar 87% adalah benar. Hasil *recall* adalah 0.84, yang berarti model berhasil mendeteksi sekitar 84% dari semua contoh kelas yang sebenarnya ada dalam dataset. *F1-score*: Ini adalah rata-rata harmonik antara precision dan recall. *F1-score* memberikan pandangan yang lebih seimbang antara precision dan recall, yang bermanfaat ketika kedua metrik tersebut berkonflik. Semakin tinggi nilai *f1-score* maka model semakin baik. Hasil *f1-score* adalah 0.86. *Accuracy*: Ini adalah akurasi dari model dalam memprediksi keseluruhan dataset. Dalam kasus ini, akurasi adalah 0.86, yang berarti model Anda memprediksi dengan benar sekitar 86% dari semua dalam dataset pengujian.

KESIMPULAN

Algoritma mengklasifikasi dengan benar saat hasil uji menggunakan confusion matrix menunjukkan nilai *Accuracy* 0,86 yang menunjukkan klasifikasi dataset pengujian ini sebesar 86% dan 14% tidak terklasifikasi dengan benar. Secara kesuluruhan, algoritma mengklasifikasi

dataset penyakit jantung dengan baik. Hal ini di indikasikan dengan nilai rata rata seperti precision 0,87 yang berarti dari semua prediksi yang model dibuat sekitar 87% adalah benar, kemudian hasil *recall* adalah 0.84, yang berarti model berhasil mendeteksi sekitar 84% dari semua contoh kelas yang sebenarnya ada dalam dataset dan hasil *f1-score* adalah 0.86 yang berarti *f1-score* bernilai 86%, semakin tinggi nilai *f1-score* maka model semakin baik.

DAFTAR PUSTAKA

[1] A. Prasetio, M. H. Hasibuan, and P. Sitompul, "Simulasi Penerapan Metode Decision Tree (C4.5) Pada Penentuan Status Gizi Balita," *Jurnal Nasional Komputasi dan Teknologi Informasi*, vol. 4, no. 3, 2021.

[2] R. Annisa, "ANALISIS KOMPARASI ALGORITMA KLASIFIKASI DATA MINING UNTUK PREDIKSI PENDERITA PENYAKIT JANTUNG," *Jurnal Teknik Informatika Kaputama (JTIK)*, vol. 3, no. 1, 2019.

[3] A. Riani, Y. Susianto, and N. Rahman, "Implementasi Data Mining Untuk Memprediksi Penyakit Jantung Mengunakan Metode Naive Bayes," *Journal of Innovation Information Technology and Application (JINITA)*, vol. 1, no. 01, pp. 25–34, Dec. 2019, doi: 10.35970/jinita.v1i01.64.

[4] "Medical gallery of Blausen Medical 2014," *WikiJournal of Medicine*, vol. 1, no. 2, 2014, doi: 10.15347/wjm/2014.010.

[5] J. Han *et al.*, "Designing Data-Intensive Web Applications."

[6] A. H. Nasrullah, "IMPLEMENTASI ALGORITMA DECISION TREE UNTUK KLASIFIKASI PRODUK LARIS," vol. 7, no. 2, 2021, [Online]. Available: http://ejournal.fikom-unasman.ac.id

[7] L. Navia *et al.*, "Klasifikasi Nasabah Menggunakan Algoritma C4.5 Sebagai Dasar Pemberian Kredit," vol. 1, no. 2, 2016.

[8] T. Indriyani *et al.*,"An Improve KNN Method for Classification of Sexually Transmitted Diseases_Rev". *International Conference on Vocation Educationand Electrical Engineering (ICVEE).* 2023 IEEE

[9] T. Indriyani, M. I. Utoyo, and R. Rulaningtyas, "A New Watershed Algorithm for Pothole Image Segmentation," *Studies in Informatics and Control*, vol. 30, no. 3, pp. 131–139, 2021, doi: 10.24846/v30i3y202112.

[10] T. Indriyani, S. Nurmuslimah, A. Taufiqurrahman, R. K. Hapsari, C. N. Prabiantissa, and A. Rachmad, "Steganography on Color Images Using Least Significant Bit (LSB) Method," 2023, pp. 39–48. doi: 10.2991/978-94-6463-174-6_5.